Published online 18 November 2009 *Nucleic Acids Research, 2010, Vol. 38, Database issue* ***D765–D773***
doi:10.1093/nar/gkp852

BioDrugScreen: a computational drug design resource for ranking molecules docked to the human proteome

Liwei Li[1,2], Khuchtumur Bum-Erdene[1,2], Peter H. Baenziger[1,2], Joshua J. Rosen[3], Jamison R. Hemmert[1,2], Joy A. Nellis[1,2], Marlon E. Pierce[3] and Samy O. Meroueh[1,2,*]

[1]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA, [2]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA and [3]Community Grids Laboratory of the Pervasive Technology Institute, Indiana University, Bloomington, IN, USA

Received August 29, 2009; Revised and Accepted September 23, 2009

ABSTRACT

BioDrugScreen **is a resource for ranking molecules docked against a large number of targets in the human proteome. Nearly 1600 molecules from the freely available NCI diversity set were docked onto 1926 cavities identified on 1589 human targets resulting in >3 million receptor–ligand complexes requiring >200 000 cpu-hours on the TeraGrid. The targets in *BioDrugScreen* originated from Human Cancer Protein Interaction Network, which we have updated, as well as the Human Druggable Proteome, which we have created for the purpose of this effort. This makes the *BioDrugScreen* resource highly valuable in drug discovery. The receptor–ligand complexes within the database can be ranked using standard and well-established scoring functions like AutoDock, DockScore, ChemScore, X-Score, GoldScore, DFIRE and PMF. In addition, we have scored the complexes with more intensive GBSA and PBSA approaches requiring an additional 120 000 cpu-hours on the TeraGrid. We constructed a simple interface to enable users to view top-ranking molecules and access purchasing and other information for further experimental exploration.**

INTRODUCTION

Virtual screening of large chemical libraries has become a widely used tool to search for small molecules that bind to a receptor of interest (1). A number of success stories have been reported in the literature (2–4). The process consists of three steps: docking, scoring and ranking. Docking is the series of computational procedures to predict the 3D structure of a receptor–ligand complex. It consists of the sampling of large number of conformations of the ligand until the most energetically favorable complex is found (5). This process is driven by sophisticated algorithms such as simulated annealing and the genetic algorithm (6). One of the first computer programs to implement docking was DOCK, written by Irwin Kuntz and coworkers (7) at the University of California San Francisco. Subsequently other programs, such as AutoDock, became widely used (8). Today there exists a number of programs in addition to AutoDock, such as Gold (9), FlexX (10), Glide (11) and others.

In virtual screening, each member of the chemical database is docked onto the receptor, resulting in as many receptor–ligand complexes as the number of molecules contained in the chemical library. The challenge next is to exploit this structural information to identify active compounds. This is accomplished through the computational prediction of the binding affinity of a compound to the receptor in a step known as scoring. A score is a number that approximates the free energy of binding of a compound to its receptor. Hence, the lower the score, the more potent is the compound. A score is assigned to all members of the docked complexes and is used to rank the docked molecules. The top compounds can then be purchased or chemically synthesized to be tested in the laboratory. While significant inroads have been made with docking, scoring remains challenging as evidenced by the large number of scoring functions that have been developed over the years; these include empirical (12–14), knowledge based (15–18) or physics based (7,19–21).

Despite its usefulness, virtual screening remains out of reach to a large number of scientists who could benefit significantly from this technique. Molecules that emerge from virtual screening could either serve as leads in drug discovery or as molecular probes in chemical biology

*To whom correspondence should be addressed. Tel: +1 317 274 8315; Fax: +1 317 278 9217; Email: smeroueh@iupui.edu

efforts. The main impediment to the wide use of virtual screening is the computational expertise required to perform the calculations as well as the significant computational resources needed to perform the calculations. To remedy this situation, we created a Web portal known as *BioDrugScreen* where we have docked 1592 molecules to nearly 2000 cavities within 1589 targets of relevance to cancer and other diseases. The docked molecules can be ranked with an array of nine scoring functions. The cancer targets originate from a recently created database known as the Human Cancer Protein Interaction Network (HCPIN). We also considered human targets for drugs approved by the Food and Drug Administration (FDA). In this effort, all cavities that can accommodate small molecules were included. This makes it possible for scientists to target alternative allosteric sites on their receptor of interest. We have created a simple interface within *BioDrugScreen* to enable users to access the ranking of molecules that were docked to their target of interest. It is worth mentioning other efforts that have worked to facilitate the search of small molecules without explicit docking, but through a pharmacophore approach (22).

MATERIALS AND METHODS

3D structures of protein targets

Protein targets included in the database are obtained from HCPIN (23) and DrugBank v2.5 (24,25) databases. The first HCPIN release contained structures up to February of 2006. We worked to create a local updated version of the database that we call HCPIN-2009. We collected sequence information of all HCPIN targets at the UniProt Web site (http://www.uniprot.org) using the SwissProt name provided by the HCPIN Web site. Targets without a SwissProt name were not considered. DrugBank, on the other hand, provides sequence information of all known targets directly at the DrugBank Web site. In total, we collected 3155 human targets, 1147 and 2241 corresponded to DrugBank and HCPIN targets, respectively. There are 233 sequences that are in common between DrugBank and HCPIN. Protein structural coverage was determined by running a BLAST (v2.2.19) protein sequence search against nonredundant Protein Data Bank (PDB) sequences (March 2009). Only PDB structures with sequence identity >80% and BLAST *E*-value $<10^{-6}$ were considered. This resulted in 1589 protein targets that have structural coverage for both HPCIN and Human Druggable Proteome (HDP), 1412 among them are human targets. Relibase+ (v2.2.2) (26) was used to identify suitable ligand binding pockets on human proteins. A total of 1926 cavities were identified on 1589 human targets with restriction of pocket volume to be within 200–4000 $Å^3$.

Docking to proteome targets

Crystal structures of human targets were obtained from RCSB database in PDB format. PDB files were processed by removing all solvents, counterions and substrates. The Reduce program (27) was used to add hydrogen atoms to proteins and optimize some of the residue orientations. The MGLTools (v1.5.2) (28) was used to assign Gasteiger charges to the protein and generate structural file for docking. The binding box was centered on the pocket identified by the Relibase+ program with side length set to 18 Å, which is large enough to encompass the entire pocket in the majority of cases. Affinity grids on the binding pocket were constructed using AutoGrid4 (8) with grid spacing of 0.375 Å. Each grid map consisted of $48 \times 48 \times 48$ grid points. The National Cancer Institute (NCI) diversity set of compounds was obtained from the ZINC8 (29) website. Compound isomers bearing the same ZINC ID were not considered, resulting in a total of 1592 compounds. Autodock4 (8) was used to dock NCI compounds to the pockets. Rigid receptor/flexible ligand protocol was used in the docking process. Compound conformational space was explored employing the Lamarckian genetic algorithm. Each docking job consisted of five runs. The maximum number of energy evaluations was set to 400 000. All other parameters were set to default values.

Scoring of the docked complexes

In addition to the AutoDock score, which was computed during the docking process, protein–ligand complexes were scored with DFIRE (30), X-Score (31), PBSA, GBSA, DockScore, GoldScore, ChemScore and PMF. The latter four scores were computed with the CScore module in SYBYL (Tripos Inc., St Louis, MO, USA). The PBSA and GBSA were calculated with the AMBER9 package (32), following similar procedure described elsewhere (33). The protein and ligand interaction was modeled using AMBER9 force field with 'parm99' and 'gaff' parameters. Ligands were assigned AM1-BCC charges (34,35) with the ANTECHAMBER program, which is part of the AMBER package. The protein–ligand complexes were first subjected to 1500 steps of in vacuum energy minimization followed by 50 steps of implicit solvent energy minimization. The MM-PBSA Perl script from the AMBER package was used to compute the binding energies. It involves molecular mechanics, Poisson–Boltzmann (PB) electrostatics (36), Generalized Born (GB) model (37) and solvent accessible surface area (SASA) calculations.

RESULTS

Protein targets in *BioDrugScreen*

In our initial effort toward our long-term objective of docking the entire human proteome, we considered two sets of proteins. The first set includes cancer targets that are obtained from a recently created dataset known as the HCPIN. It contains proteins from seven different signaling pathways that are associated with cancer together with their interactions. The unique aspect of HCPIN is that 3D structures are assigned to all proteins within the dataset. The structures fall into two categories. The first is a core set of structures known as 'pathway proteins' that are directly implicated in cancer. These proteins are obtained directly from the Kyoto Encyclopedia of Genes

and Genomes (KEGG) database. Other proteins not in KEGG but interact with the core proteins are known as 'interaction proteins'. A web site was created to provide information about the interaction network that can be found at http://nesg.org:9090/HCPIN/index.jsp. In an effort to update the structural coverage of the HCPIN database, we collected the latest sequence information of HPCIN targets and searched the PDB for structures that were released since 2006. We refer to this local updated version as human cancer proteome database HCPIN-2009. This was done for both 'pathway' and 'interaction' proteins resulting in a total of 1203 targets with structures.

In *BioDrugScreen*, we also consider proteins that are targets of existing FDA-approved drugs. Our interest in these proteins stemmed not only from their established druggable potential, but also from the fact that they can serve to develop drugs for the treatment of diseases other than cancer. In addition, they can be used as 'normal' proteins during a search for cancer therapeutics. The 3D structure of these proteins were retrieved from the DrugBank database (38), as described in the 'Materials and Methods' section. This is an annotated resource that combines detailed drug data with comprehensive drug target and drug action information. Among all the sequences that we obtained from DrugBank v2.5, we identified 1147 unique human drug targets. It was found that 530 of them possessed a corresponding PDB crystal structure.

Mapping binding cavities of targets in *BioDrugScreen*

Receptor–ligand interactions are driven by a complex array of forces that include polar, nonpolar and entropy (39). A requirement for optimal binding is the presence of a cavity within the receptor that can accommodate a small molecule ligand. For example, most enzymes possess well-defined cavities that have been fine-tuned during evolution to exquisitely accommodate a substrate (39). But when the ligand is a protein, these cavities are no longer prerequisite, since the protein–protein interaction occurs over a large surface that can be shielded from solvent to promote favorable nonpolar interactions (40,41). While it is challenging to design small molecules that target such surfaces, distal sites may be exploited to modulate the protein–protein interactions. Therefore, as we seek to map binding sites within the human cancer and druggable proteomes, we considered all cavities that can accommodate small molecules, whether or not they occur within an enzyme active site or protein-interaction site.

In virtual screening, molecular docking on the entire protein surface is possible, but computationally prohibitive. It is typical that docking is focused on a cavity of interest within a protein. For most enzymes, this cavity is usually the active site, as identified by biochemical studies or a crystal structure of the enzyme–substrate complex. In most other cases, the cavity must be identified by scanning the entire surface of the protein for depressions and crevasses that can accommodate small molecules. A number of computational methods have been developed for that purpose over the years. In this work, we following the approach implemented within the Relibase+ program (26) for identifying cavities within a protein surface reported in literatures (42,43). The process is mostly driven by an external computer program LigSite (44) to scan the surface of the proteins within the PDB databank.

Docking the human proteome

The identification of cavities within the HCPIN and HDP set the stage to dock small molecules to the targets within these databases. To perform the docking, the AutoDock 4 program (45) requires the construction of an affinity grid around each cavity. This pre-computed grid significantly reduces computing costs during docking. Pair potentials between atoms in a ligand and the receptor are computed based on the position of the ligand atom within the grid during the docking. The pair-wise potential is determined by identifying the value within a table, which is significantly faster than computing the value at each step. The construction of the grids for all the cavities within *BioDrugScreen* was performed using an in house python script using the center of the binding cavity as described in the 'Materials and Methods' section. We estimate that the computation of the Grid for all pockets within the receptors considered in this work was ~70 cpu-hours.

Once the grids were computed for all the cavities within *BioDrugScreen*, we initiated the docking of small molecules. A set of 1592 compounds from the NCI diversity set were prepared for docking as described in the 'Materials and Methods' section. These molecules were docked to each of the cavities in *BioDrugScreen*. Hence a total of 1926 cavities × 1592 molecules = 3 066 192 complexes resulted from this initiative. The coordinates for the binding pose of each of these complexes is provided for download at the *BioDrugScreen* site. It is estimated that the docking effort required an equivalent of ~200 000 h of computer time.

Scoring the docked human proteome

A key step toward identifying active molecules in virtual screening is scoring, which is a process that ranks molecules based on binding affinity. Over the years, a number of scoring functions have been developed. It is well known that the performance of these scoring functions is highly dependent on the receptor. To that end, instead of relying on a single scoring function, we have scored all docked structures in *BioDrugScreen* with widely used established scoring functions. We parenthetically add that *BioDrugScreen* also includes an option for users to create their own scoring function. But this is an optional step, as ranking of molecules can be accomplished with widely used scoring functions that we have pre-computed for the complexes in the database. They include knowledge-based statistical potentials such as the DFIRE and PMF scoring functions. Empirical functions such as AutoDock, GoldScore, X-Score and ChemScore are also included. Unlike the knowledge-based potentials, these scoring functions are developed using experimental structural and ligand affinity data. Finally, DockScore is a force field-based scoring function that consists of a

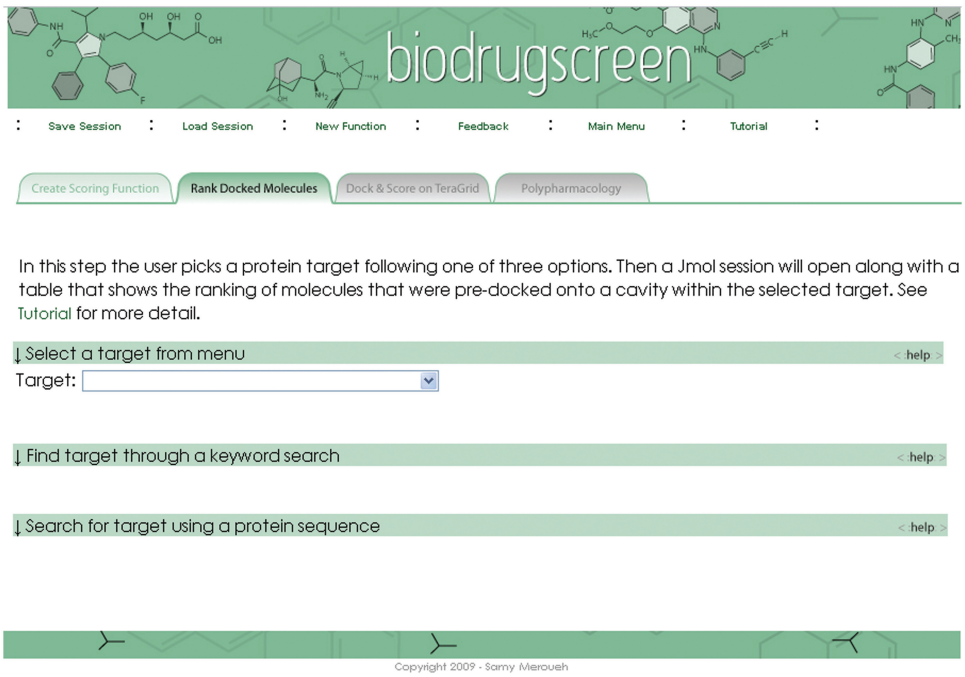

Figure 1. A snapshot of the *BioDrugScreen* page with options to rank ligands that were pre-docked to a large number of cavities of receptors from the human proteome.

Lennard Jones and Coulomb potential summed over all pair-wise interactions between ligand and receptor.

In addition to standard scoring functions, we have used the force field-based PBSA and GBSA approaches to score receptor–ligand complexes (33). These physics-based methods are significantly more intensive in their computational requirements than standard scoring functions. They are similar to the widely used MM–PBSA method, except that the calculations are performed on a single energy minimized structure of docked ligand/target complex. The score consists of potential energy determined by summing of all pair-wise energies using a Lennard Jones and Coulomb potentials. In addition, a solvation term is added that consists of polar and nonpolar contributions. The nonpolar solvation corresponds to the cost associated with creating a cavity in the solvent. It is determined by computing the solvent-accessible surface area of the molecule. The polar term consists of a PBSA or Generalized-Born (GBSA) calculations that measures the effect of introducing a charged cavity within a solvent.

Ranking pre-docked receptor–ligand complexes in *BioDrugScreen*

Upon accessing the http://www.biodrugscreen.org main page, the user is presented with introductory text describing some of the features within the site. Below the text, the user clicks on the 'Start' button to get access to the pre-docked database. A page appears with two options shown as 'RANK' and 'DERIVE' links. The 'DERIVE' option is an advanced feature within *BioDrugScreen* that enables users to create their own customized scoring function and apply it to rank receptor–ligand complexes within the portal. However, this feature is not discussed in this manuscript, as it relates to the server aspect of *BioDrugScreen*. The other option is for the ranking pre-docked receptor–ligand complexes.

When the user clicks on the 'RANK' button, they are taken to a page to select their target of interest and get started with ranking molecules. In this page, which is shown in Figure 1, the user will find three options to select a target. In the first option (Figure 2A), the user clicks on a pull-down menu to scroll down a list of the protein targets within *BioDrugScreen*. The list is ordered in alphabetical order using the name of the target that is obtained from the RCSB web site under 'Molecular Description'. Once a target is selected, another pull-down menu will appear next to it prompting the user to select a chain. The chains correspond to different binding partners within a protein–protein complex. More than one chain will appear only if the crystal structure involves more than one protein. Once a chain is selected, a third pull-down menu appears that lists the number of cavities. Once a cavity is selected, a Jmol session will open. The session depicts the 3D structure of the receptor in wire representation and a surface is constructed around the cavity to facilitate viewing of the shape of the cavity. The user can find instructions to display the receptor in surface representation to facilitate viewing of the binding cavity of the docked molecules.

The second option for selecting a target is shown in Figure 2B. The user can perform a keyword search for a target of interest. Once the 'Search' button is clicked, a table will appear with a list of receptors that matched the search criterion. In the third column within the table, a list

of cavities within the target is provided. When the user chooses a cavity, a Jmol session will appear along with a table ranking the compounds. Finally, *BioDrugScreen* also provides a third option to identify targets through a BLAST protein sequence search as shown in Figure 2C. Once a sequence is pasted into the search box, a BLAST search is performed and a table with sequences that match the query sequence is shown. In the second column of the

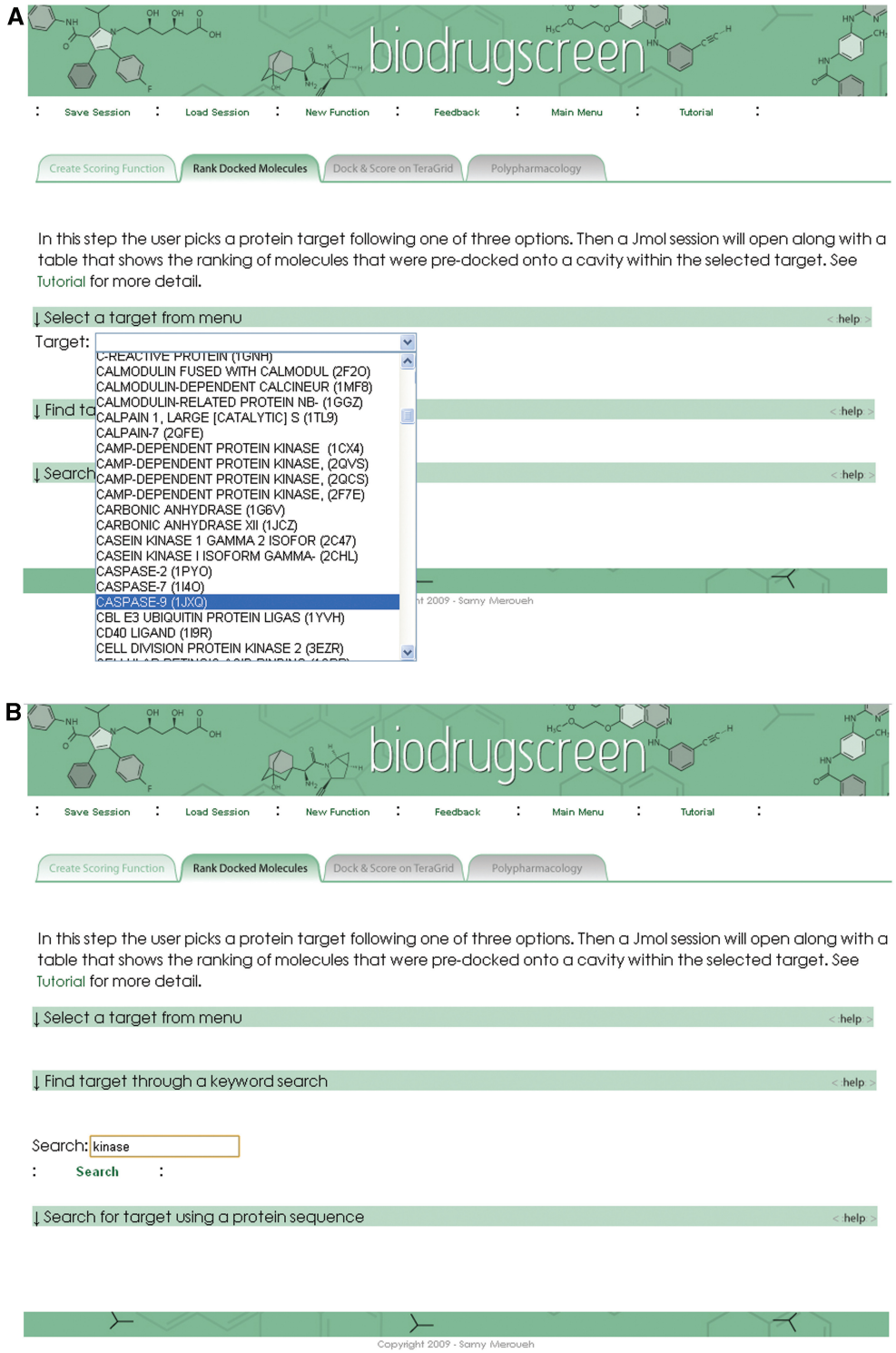

Figure 2. A snapshot depicting the three search options available in *BioDrugScreen*. (**A**) A pull-down menu lists all the targets within *BioDrugScreen*. The user then selects the 'Chain' and 'Binding Cavity' pull-down menus to select chain and cavity; (**B**) targets can be found through a keyword search. (**C**) Targets can be identified through a BLAST sequence search. The user pastes the sequence within the box provided to conduct the search.

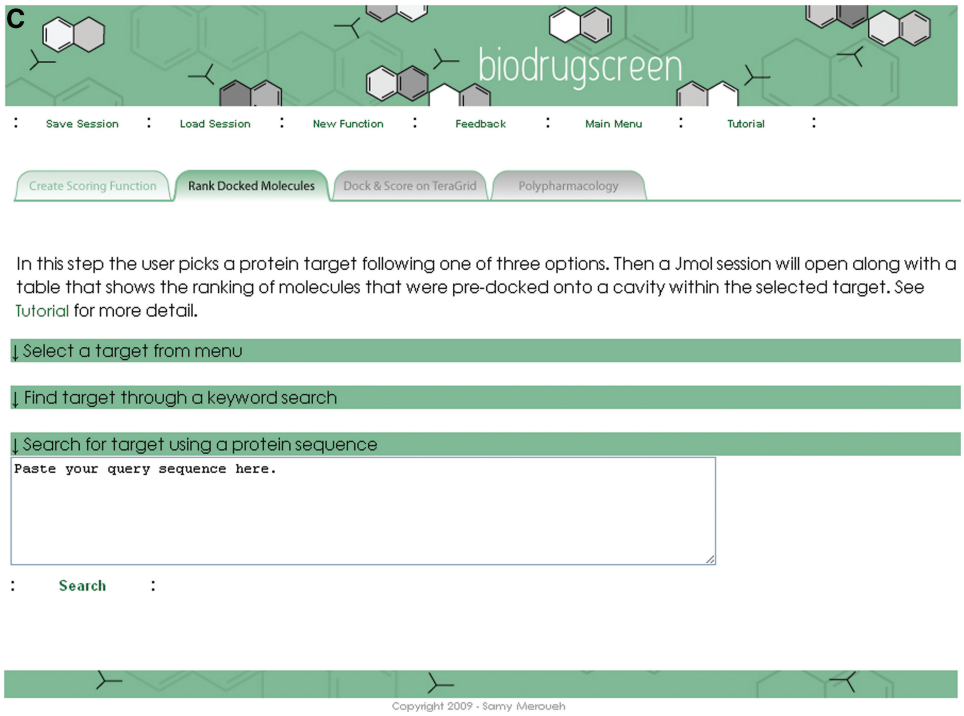

Figure 2. Continued.

table, a list of cavities for each receptor is provided. When a cavity is selected, a Jmol session along with a table with ranked compounds is shown.

Below the Jmol session, a table is created that lists the ranking of the receptor–ligand complexes of the NCI diversity set docked onto the receptor of interest. The table lists the top molecules as ranked by 9 scores, namely AutoDock, ChemScore, DockScore, DFIRE, GoldScore, PMF, X-Score, PBSA and GBSA (Fig. 3). Each row corresponds to a different molecule. But the molecules need not be the same for each scoring function. For each molecule, we provide a total of five links. The ''ZINC'' link, leads to purchasing and other information at the ZINC Web site at http://zinc.docking.org/index.shtml. This web site is a database of commercially available compounds for virtual screening (46). The 'PubChem' link leads to additional detailed information for the molecules at the PubChem (47) web site at http://pubchem.ncbi.nlm.nih.gov/. PubChem is a large public database of biological data for small molecules. It is significantly larger than the ZINC database, as it is not strictly limited to commercially available molecules. The 'MOL2' icon leads to coordinates of the docked molecule in mol2 format. Users can also download the coordinates of the receptor with a link located at the bottom of the Jmol session. These coordinates are used to depict the small molecule bound to the receptor in the Jmol session. The 'COMPARE' link provides the ranking of each molecule in other scoring functions. Finally, clicking on the 'JMOL' button will take a user to the Jmol session that shows the 3D structure of the ligand bound to receptor. The receptor is shown in line representation, while the bound molecule is depicted in capped-sticks. Both receptor and ligand are color-coded based on atom types (C, N and O are shown in gray, blue and red, respectively). Each page lists the top 50 molecules. The remaining molecules can be accessed by clicking on the numbers at the top of the table ranging from 1 to 32.

DISCUSSION

Virtual screening is now widely used to search for small molecules ligands that bind to macromolecules (2–4). The method is used extensively in drug discovery efforts. However, the use of virtual screening has been limited to individuals and research groups with significant computational expertise and resources. We seek to remedy this situation by creating a web resource known as *BioDrugScreen* that contains about 3 000 000 receptor complexes that were obtained by docking about 1600 compounds from the NCI diversity set to nearly 2000 cavities within targets of the human proteome. These complexes were scored with established scoring functions, such as AutoDock, PMF, DFIRE, ChemScore and GoldScore. We also performed continuum electrostatic calculations on all these complexes using the PB and Generalized-Born methods and combined with results with intermolecular potential energy from the Amber force field. The resulting GBSA and PBSA terms are also provided for ranking receptor–ligand complexes within *BioDrugScreen*.

Our expectation is that the 'Rank Ligand' step within *BioDrugScreen* will be used by users with a wide range of computational expertise. It simply requires the selection of a target from one of three possible options. A table is then

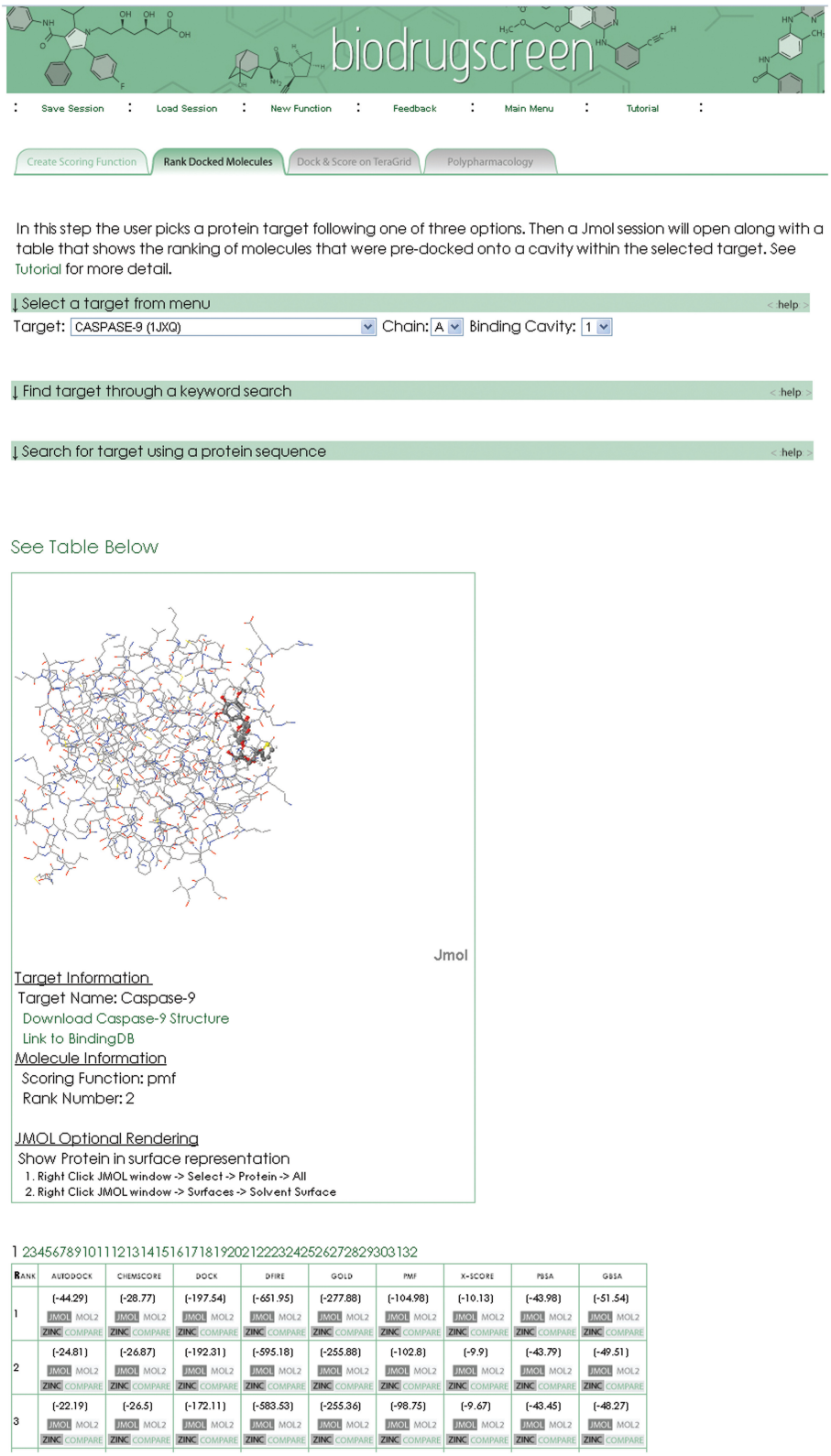

Figure 3. Snapshot of *BioDrugScreen* following the selection of a receptor through one of the three mechanisms provided. A Jmol session depicts the 3D structure of a molecule bound to the protein Caspase-9. The molecule is shown in capped-sticks representation with atoms color-coded based on atom types (N, C, S and H in blue, grey, yellow and white, respectively). The protein is shown in wire rendering. A table below the Jmol session provides a listing of the molecules docked to the target and ranked using a series of standard scoring functions along with various links for further structural analysis and information from other sites.

created showing the ranking of the molecules from the NCI dataset that are docked onto the target. The molecules are ranked from best to worse, such that the higher the rank, the greater the likelihood that the molecules will bind to the target. Users can analyze the 3D structure of the complexes and access purchasing information for these molecules directly from the ZINC web site through the various links provided for each entry (molecule) in the table. Since a large number of the targets within *BioDrugScreen* are of relevance to cancer, it is our expectation that the database will serve to identify new anticancer agents.

While this manuscript was focused on the significant receptor–ligand database within *BioDrugScreen*, it is worth mentioning that the web portal provides additional options for users with computational expertise. We envision that these options will significantly enhance the database aspect of the portal. The first such option enables users to create and validate their own scoring functions and to apply these custom scoring functions to rank the docked chemical libraries within the database. The user can create a training set using structural and affinity data from PDBcal, PDBbind and BindingDB. Subsequently, the user can choose the components that will make up their scoring functions from among more than 20 pre-computed descriptors. A regression analysis is then performed to derive the empirical function, which can be validated through enrichment and ROC plots on 40 targets that have been pre-docked with active and decoy molecules from the Directory of Useful Decoys validation set. The custom scoring function can be used to rank docked molecules. *BioDrugScreen* incorporates another sophisticated option that lets users dock molecules to their favorable target on the TeraGrid and score them with GBSA by simply uploading their PDB file. *BioDrugScreen* includes a feature that lets users monitor their job until it is completed. All user-initiated docked complexes are uploaded automatically into the database but will only be visible to the user who docked the molecules.

FUNDING

Lungs for Life Lung Cancer Working Group Fellowship (LL); National Institutes of Health; Lilly Endowment, Inc. (INGEN grant); IBM, Inc. (computer time on the Big Red supercomputer at Indiana University is funded by the National Science Foundation as well as by Shared University Research grants to Indiana University); Indiana METACyt Initiative; Lilly Endowment, Inc. (partially by Indiana METACyt Initiative of Indiana University). Funding for open access charge: Lilly Endowment, Inc. (INGEN grant).

Conflict of interest statement. None declared.

REFERENCES

1. Shoichet,B.K. (2004) Virtual screening of chemical libraries. *Nature*, **432**, 862–865.
2. Song,H., Wang,R., Wang,S. and Lin,J. (2005) A low-molecular-weight compound discovered through virtual database screening inhibits Stat3 function in breast cancer cells. *Proc. Natl Acad. Sci. USA*, **102**, 4700–4705.
3. Bowman,A.L., Nikolovska-Coleska,Z., Zhong,H., Wang,S. and Carlson,H.A. (2007) Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models. *J. Am. Chem. Soc.*, **129**, 12809–12814.
4. Betzi,S., Restouin,A., Opi,S., Arold,S.T., Parrot,I., Guerlesquin,F., Morelli,X. and Collette,Y. (2007) Protein protein interaction inhibition (2P2I) combining high throughput and virtual screening: application to the HIV-1 Nef protein. *Proc. Natl Acad. Sci. USA*, **104**, 19256–19261.
5. Kitchen,D.B., Decornez,H., Furr,J.R. and Bajorath,J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev.*, **3**, 935–949.
6. Dias,R. and de Azevedo,W.F. (2008) Molecular docking algorithms. *Curr. Drug Targets*, **9**, 1040–1047.
7. Meng,E.C., Shoichet,B.K. and Kuntz,I.D. (1992) Automated docking with grid-based energy evaluation. *J. Comput. Chem.*, **13**, 505–524.
8. Morris,G.M., Goodsell,D.S., Halliday,R.S., Huey,R., Hart,W.E., Belew,R.K. and Olson,A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, **19**, 1639–1662.
9. Berendsen,H.J.C., Vanderspoel,D. and Vandrunen,R. (1995) Gromacs – a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.*, **91**, 43–56.
10. Rarey,M., Kramer,B., Lengauer,T. and Klebe,G. (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, **261**, 470–489.
11. Friesner,R.A., Banks,J.L., Murphy,R.B., Halgren,T.A., Klicic,J.J., Mainz,D.T., Repasky,M.P., Knoll,E.H., Shelley,M., Perry,J.K. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **47**, 1739–1749.
12. Bohm,H.J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.*, **8**, 243–256.
13. Eldridge,M.D., Murray,C.W., Auton,T.R., Paolini,G.V. and Mee,R.P. (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.*, **11**, 425–445.
14. Bohm,H.J. (1998) Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.*, **12**, 309–323.
15. Muegge,I. and Martin,Y.C. (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.*, **42**, 791–804.
16. Muegge,I. (2006) PMF scoring revisited. *J. Med. Chem.*, **49**, 5895–5902.
17. Gohlke,H., Hendlich,M. and Klebe,G. (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, **295**, 337–356.
18. Velec,H.F., Gohlke,H. and Klebe,G. (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.*, **48**, 6296–6303.
19. Huey,R., Morris,G.M., Olson,A.J. and Goodsell,D.S. (2007) A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.*, **28**, 1145–1152.
20. Jones,G., Willett,P. and Glen,R.C. (1995) A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput. Aided Mol. Des.*, **9**, 532–549.
21. Jones,G., Willett,P., Glen,R.C., Leach,A.R. and Taylor,R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **267**, 727–748.
22. Schneidman-Duhovny,D., Dror,O., Inbar,Y., Nussinov,R. and Wolfson,H.J. (2008) PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Res.*, **36**, W223–W228.

23. Huang,Y.J., Hang,D., Lu,L.J., Tong,L., Gerstein,M.B. and Montelione,G.T. (2008) Targeting the human cancer pathway protein interaction network by structural genomics. *Mol. Cell. Proteomics*, **7**, 2048–2060.
24. Wishart,D.S., Knox,C., Guo,A.C., Cheng,D., Shrivastava,S., Tzur,D., Gautam,B. and Hassanali,M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
25. Wishart,D.S., Knox,C., Guo,A.C., Shrivastava,S., Hassanali,M., Stothard,P., Chang,Z. and Woolsey,J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
26. Hendlich,M., Bergner,A., Gunther,J. and Klebe,G. (2003) Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.*, **326**, 607–620.
27. Word,J.M., Lovell,S.C., Richardson,J.S. and Richardson,D.C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, **285**, 1735–1747.
28. Sanner,M.F. (1999) Python: a programming language for software integration and development. *J. Mol. Graphics Mod.*, **17**, 57–61.
29. Irwin,J.J. and Shoichet,B.K. (2005) ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model*, **45**, 177–182.
30. Zhang,C., Liu,S., Zhu,Q. and Zhou,Y. (2005) A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes. *J. Med. Chem.*, **48**, 2325–2335.
31. Wang,R., Lai,L. and Wang,S. (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.*, **16**, 11–26.
32. Case,D.A., Cheatham,T., Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
33. Li,L., Uversky,V.N., Dunker,A.K. and Meroueh,S.O. (2007) A computational investigation of allostery in the catabolite activator protein. *J. Am. Chem. Soc.*, **129**, 15668–15676.
34. Jakalian,A., Bush,B.L., Jack,D.B. and Bayly,C.I. (2000) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.*, **21**, 132–146.
35. Jakalian,A., Jack,D.B. and Bayly,C.I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, **23**, 1623–1641.
36. Fogolari,F., Brigo,A. and Molinari,H. (2002) The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.*, **15**, 377–392.
37. Still,W.C., Tempczyk,A., Hawley,R.C. and Hendrickson,T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, **112**, 6127–6129.
38. Wishart,D.S., Knox,C., Guo,A.C., Cheng,D., Shrivastava,S., Tzur,D., Gautam,B. and Hassanali,M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
39. Li,L. and Meroueh,S.O. (2008) *Encyclopedia for the Life Sciences*. John Wiley and Sons, London.
40. Liang,S., Li,L., Hsu,W.L., Pilcher,M.N., Uversky,V., Zhou,Y., Dunker,A.K. and Meroueh,S.O. (2009) Exploring the molecular design of protein interaction sites with molecular dynamics simulations and free energy calculations. *Biochemistry*, **48**, 399–414.
41. Li,L., Liang,S., Pilcher,M.M. and Meroueh,S.O. (2009) Incorporating receptor flexibility in the molecular design of protein interfaces. *Protein Eng. Des. Sel.*, **22**, 575–586.
42. Kuhn,D., Weskamp,N., Schmitt,S., Hullermeier,E. and Klebe,G. (2006) From the similarity analysis of protein cavities to the functional classification of protein families using Cavbase. *J. Mol. Biol.*, **359**, 1023–1044.
43. Kuhn,D., Weskamp,N., Hullermeier,E. and Klebe,G. (2007) Functional classification of protein kinase binding sites using cavbase. *ChemMedChem*, **2**, 1432–1447.
44. Huang,B.D. and Schroeder,M. (2006) LIGSITE(csc): predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.*, **6**, 19.
45. Morris,G.M., Huey,R. and Olson,A.J. (2008) Using AutoDock for ligand-receptor docking. *Curr. Protoc. Bioinformatics*, Chapter 8, Unit 8.14.
46. Irwin,J.J. and Shoichet,B.K. (2005) ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model*, **45**, 177–182.
47. Wang,Y.L., Xiao,J.W., Suzek,T.O., Zhang,J., Wang,J.Y. and Bryant,S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, **37**, W623–W633.